

WOLFGANG VON KEMPELEN'S 'SPEAKING MACHINE' AS AN INSTRUMENT FOR DEMONSTRATION AND RESEARCH

Jürgen Trouvain¹ & Fabian Brackhane²

¹Institute of Phonetics, Saarland University, Saarbrücken, Germany

²Institut für deutsche Sprache (IDS), Mannheim, Germany

trouvain@coli.uni-saarland.de | brackhane@pragmatik.ids-mannheim.de

ABSTRACT

Scientific interest in von Kempelen's 'speaking machine' stems mainly from a general interest in the history of science. This study, however, is devoted to the question of what relevance the 'speaking machine' has today. Apart for discussing why it fascinates researchers and non-researchers alike we describe the potential of replicas as an instrument for demonstration and for researching speech generation.

Keywords: von Kempelen, history of phonetics, mechanical speech synthesis.

1. INTRODUCTION

Wolfgang von Kempelen's 1791 book "Mechanismus der menschlichen Sprache" ("Mechanism of Human Speech and Language") [19] and the description of his 'speaking machine' therein have great historical relevance for the phonetic sciences (cp. e.g. [4, 9, 14]). Various replicas of the 'speaking machine' are witness to its popularity and its unique position in research dedicated to speech generation [11, 18]. Scientific interest in von Kempelen's 'speaking machine' stems mainly from a general interest in the history of science. This study is devoted to the question of what relevance the 'speaking machine' has today.

2. THE 'SPEAKING MACHINE' AS AN INSTRUMENT FOR DEMONSTRATION

The 'speaking machine' has always been found to be an extra-ordinary and convincing instrument for demonstrations (see Figure 1). This is true for the European courts in von Kempelen's time as well as for today's classrooms. An instrument just consisting of wood, metal, leather, rubber and a bit of ivory has a fascinating effect despite, or perhaps because of the electronic methods of generating speech that have been in use now for several decades.

Even if we try to explain the fascination with the 'speaking machine' by pointing out its authenticity and simplicity as well as its reproducibility [17], it is important to identify the people who are attracted to it, what they find attractive and what their scientific interests are.

Figure 1: The 'speaking machine' (inner life): palm of the hand on the right forming vowel resonances in front of a rubber funnel ("vocal tract"); the hand on the wooden windchest ("thorax") regulating nasal cavity resonances; (invisible) elbow providing pressure on the bellows ("lungs"). The reed pipe ("glottis") is located within the "nose".



After many performances with the Saarbrücken replicas [2] for very different audiences – even if the speech was usually only 'Mama' and 'Papa' – we can report that nobody was left unimpressed. This is true for students as it is for professors, for children as well as for older persons, for those with a more technical background such as engineers as well as for those with a more human than technical interest such as speech pathologists. Even those trained in the phonetic sciences are unable to resist the fascination of the 'speaking machine'.

We claim that replicas of the speaking machine can very well serve to illustrate how speech sounds are generated – in more than one modality, in fact, since the user can *see* and *touch* the machine as well as hear it – see also the do-it-yourself vowel resonators in [7]. Experiencing and understanding in multiple modalities provide an outstanding and a rather unusual opportunity to observe the process

of speaking, which is mostly invisible, unconscious and obscured by the focus on the content of what is said. One interesting aspect of demonstrations is the fact that the player of the instrument feels impelled to silently articulate in synchrony with the 'manual articulation'. Apparently it is easier for the player to articulate manually when the cognitive control of the speech articulators takes place. Possibly this 'inner speech' can be suppressed only by a conscious effort. A side effect is that the spectator has the impression that the player is articulating with the voice of the machine.

Ultimately, experiencing the 'speaking machine' prompts the question 'How can it be that this construction of wood, leather and metal can speak like a human being?', and that question inevitably leads to the core of the phonetic sciences: 'How is it that humans are able to speak?'

Von Kempelen too started in the 18th century with this problem. He was fully aware of the limitations to the practical application of his research. He finished his chief work [19] *inter alia* with the wish that his readers "give some attention to this new invention, which is still in its infancy, and that they advance it by their thinking and effort."¹

3. THE 'SPEAKING MACHINE' AS AN INSTRUMENT FOR RESEARCH

In our view, the significance of the 'speaking machine' goes beyond that of a unique instrument for demonstrating the generation of speech. We present here a few research questions regarding the speech production by humans and by machines.

3.1. Role of the sub-glottal resonance cavity

We performed tests with wind boxes of different sizes linking the reed pipe ('lingual pipe') as the phonatory element and the bellows as the 'lung'. It has been shown that the size of the wind box as the 'sub-glottal' resonant cavity has a great impact on the degree of authenticity of an artificial children's voice [2]. The results of the authenticity tests are different depending on the size of the box and the type of wood.

So the speech-production question to be answered here is what role do the sub-glottal cavity and the generated air pressure play in human speech as well as in the individual character of a person's voice. References to sub-glottal resonance features are not (yet) found in phonetic text books. There is a need for more basic research in this respect (e.g. [20]).

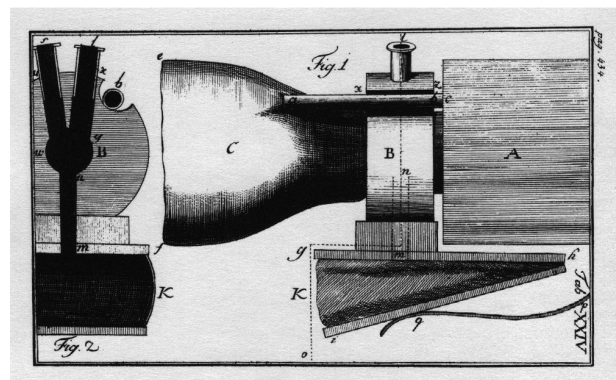
The question can be extended to voices generated by articulatory synthesis, which often still sound unnatural nowadays. One advantage of experimenting with a replica is the effortless and quick exchange of different 'sub-glottal' cavities.

3.2. Compliance with voiceless stops

One of the prerequisites for phonation is a sufficient transglottal air pressure drop to maintain an airflow. During oral closure, supra-glottal air pressure increases and leads to a reduction of the transglottal air pressure difference – and hence to devoicing. Typically voicing ceases after 15 ms [13]. The closure phases of fully voiced plosives are usually considerably longer than that and it is assumed that the vocal tract is enlarged in order to maintain a trans-glottal flow and delay the cessation of voicing.

An actively controlled enhancement of the vocal tract can be achieved for example by lowering the larynx or by lowering the tongue body. A passive enlargement of the vocal tract happens through the compliance of tissue [13]. Exactly this effect can be obtained with the 'speaking machine' by means of the 'plosive bellows'. These bellows are located directly beneath the 'nasal cavity' and the two cavities are linked with a small tube (Figure 2).

Figure 2: Plosive bellows "K" seen in cross-section in the front view (left-hand side) and in the side view of the inner life of the speaking machine (taken from the original engraving in [19]). The 'nasal cavity' ("B") is linked with a small tube ("n") to the plosive bellows located beneath.



These second, much smaller bellows also have the effect of lengthening the voiced phase of the plosives. There is evidence of this effect for various replicas. However, replicas *without* these bellows are better at generating voiceless plosives with an aspiration phase before voicing begins. An aspiration effect is not possible in replicas *with* these

bellows, where ['papa] sounds more like ['baba]. The possibility of switching the second bellows on and off would clearly offer a good solution. It must be noted, however, that it was von Kempelen's belief that he was invoking exactly the opposite effect with the installation of the additional bellows: "In order to strengthen the explosion of the unvoiced consonants I have made another equally important addition. I have attached small bellows [...]." [19: p. 437]

In order to obtain a distinction between [b] and [p] with the 'speaking machine', air pressure is increased for [p] compared to [b] immediately before and during the closure of both apertures. However, the initiation of 'sub-glottal' pressure must not be too strong, otherwise the onset of the 'vocal fold' vibration fails. (Possibly this effect applies to human crying and screaming too, but there it may be offset by adjustment of the vocal fold tension.)

3.3. The prize questions of the St. Petersburg Academy

Further thoughts concern the connection between the generation of human-like voices and organ building as it was stimulated in the prize questions of the academy of sciences in St. Petersburg in 1780 under the guidance of Leonhard Euler [10]:

"1. What is the nature and the character of the vowel letters a, e, i, o, u, which so significantly differ from each other?

2. Is it not possible to build instruments in the manner of those organ pipes which are known under the term 'vox humana' to express the sound of the vowel letters a, e, i, o, u?"

Christian Gottlieb Kratzenstein won the prize for answering these questions [10, 12]. He provided pipes which generated the requested vowels. Although his approach can be regarded as an important step towards mechanical speech synthesis [12], those pipes did not show any similarity to vowel production in a human vocal tract. Furthermore, they only generated static, isolated vowels. With the help of a sort of 'organ', an individual key for each single vowel controlled a separate pipe. In contrast, von Kempelen took an important step forward. He recognised the central role of coarticulation and built this idea into his machine [19: p. 407]:

"Now I started to understand that the single letters could be invented but, in the way I did it, never joined together in syllables, and that I had to follow nature which has only one glottis and only

one mouth out of which all sounds are emitted and only for this reason can connect with each other."

The problem of the second question, asking for the 'vox humana' remains unsolved. The term does not refer to the human voice, as it is sometimes erroneously translated (e.g. [8]), but to the organ register (or 'organ stop') which has existed in organ building for centuries (cp. [2]). This organ register is used with a so called 'tremulant' in order to generate a sound similar to the vibrato of a human singing voice. The 'tremulant' mechanism is located *before* the pipe and it steers the periodically interrupted air stream to the pipe as the instrument of excitation. A similar mechanism might be used to model machine singing voices, although human singers as well as singing synthesisers modulate glottal parameters to produce vibrato (e.g. [17, 1]).

In the course of his research von Kempelen also took up the idea of using the organ register 'vox humana' as the basis for his speech synthesiser. This is the reason why he used nothing but reed pipes, as in organ building, to act as the vocal folds (with only one vibrating element, similar to a clarinet mouthpiece).

He discarded the construction he had first developed, based on the mouthpiece of an oboe, i.e. with two elements vibrating against each other, similar to human phonation with two vibrating vocal cords. Although he knew of Kratzenstein's work, he did not follow his construction which was better in some ways (though based on a principle of phonation which was fundamentally wrong). Instead he experimented, among other things, with highly unusual modifications of organ pipes in order to achieve a sound similar to a human voice. A combination of Kratzenstein's reed pipe with von Kempelen's 'speaking machine' would be a very interesting object of research.

4. SPEECH SYNTHESIS THEN AND NOW

The 18th century could also be called century of automata, which, of course, included speech automata [9]. However, the task of these speech automata was the *rendering* of sound. Von Kempelen's invention, on the other hand, dealt with the *generation* of sound. Kempelen's speaking machine was probably the first ever functioning mechanical speech synthesiser that was able to generate short utterances. It is amazing and admirable that the historic speaking machine can stand comparison with the hardware synthesisers of the 21st century, e.g. [6]. The sound quality is

better than that of many modern ones and it is sufficient to authentically mimic a child's voice uttering a bi-syllabic word, today as in von Kempelen's time.

Originally the speaking machine was planned as an aid for the deaf. Von Kempelen recognised the strong link between speech and language competence and social acceptance: You are no one unless you can speak. This motivation can also be found for another invention of him when he developed and built a type-setting machine for a blind person [15].

In the 18th century there was not only a wish to produce synthetic speech per se. There was also a clear idea of what the synthetic voices should sound like. In 1761 Leonhard Euler wrote in his popular scientific 'Letters to a Princess' [5]:

"Without doubt it would be one of the most important discoveries to construct a machine that could properly express all sounds and tones of our speech with all articulations. [...] The preachers and orators whose voices were not strong or attractive enough could then play their sermons and discourses on such a machine, in the way that the organ players perform their pieces of music. The thing does not seem impossible to me."

At the beginning of the 21st century, 230 years after von Kempelen's invention of the 'speaking machine' we still have to face the question, whether the speech synthesisers of today are able to generate sermons and discourses as Euler envisaged. When we consider that the speech synthesis research of the last ten years has taken up topics such as emotions, affect and other forms of non-linguistic expression (cp. [16]), we can at least note some progress. But there is still a massive amount of research to do before we can give a convincing positive answer to the question. One important step is the acceptance that 'expressing speech with all articulations' means far more than the intelligible transmission of textual information.

Acknowledgments and final remark

This paper is a slightly modified and translated excerpt of [18]. The authors would like to express their gratitude to Mr. Stephan Mayer and his team (Heusweiler/Saar) for their professional support with the reconstructions. Many thanks also to Bill Barry and Eva Lasarczyk for valuable comments and helping us with the translation.

5. REFERENCES

- [1] Birkholz, P. 2007. Articulatory synthesis of singing. Special Session on "Synthesis of Singing", *Proc. Interspeech*, Antwerpen.

- [2] Brackhane, F. 2009. Die linguistische "Sprachorgel" – Über die Verbindung von Orgelbau und Sprachsynthese. *organ – Journal für die Orgel* 4/09, 42-46.
- [3] Brackhane, F. & Trouvain, J. 2008. What makes "Mama" and "Papa" acceptable? – Experiments with a replica of von Kempelen's speaking machine. *Proc. 8th Int'l Speech Production Seminar (ISSP)*, Strasbourg, 333-336.
- [4] Dudley, H. & Tarnoczy, T.H. 1950. The speaking machine of Wolfgang von Kempelen. *JASA* 22, 151-166.
- [5] Euler, L. 1761. *Lettres à une princesse d'Allemagne sur divers sujets de physique & de philosophie*. 137th letter from 16 June 1761.
- [6] Fukui, K., Ishikawa, Y., Shintaku, E., Ohno, K., Sakakibara, N., Takanishi, A. & Honda, M. 2008. Vocal cord model to control various voices for anthropomorphic talking robot. *Proc. 8th Int'l Speech Production Seminar (ISSP)*, Strasbourg, 341-344.
- [7] Huckvale, M. 2008. Make your own vowel resonators! <http://www.phon.ucl.ac.uk/home/mark/vowels/> (last visit: 04-05-2011).
- [8] Kohler, K.J. 2000. The future of phonetics. *Journal of the International Phonetic Association* 30, 1-24.
- [9] Köster, J.-P. 1972. *Historische Entwicklung von Syntheseapparaten zur Erzeugung statischer und vokalartiger Signale nebst Untersuchungen zur Synthese deutscher Vokale*. Hamburg: Buske.
- [10] Kratzenstein, Chr. G. 1781. *Tentamen Resolvendi Problema ab Akademia Scientiarum Imperiali Petropolitana ad Annum 1780 Publicae Propositum*.
- [11] Nikléczy, P. & Olaszy, G. 2003. A reconstruction of Farkas Kempelen's speaking machine. *Proc. Eurospeech*, Geneva, 2453-2456.
- [12] Ohala, J. 2011. Christian Gottlieb Kratzenstein: Pioneer in speech synthesis. *Proc. 17th ICPHS*, Hong Kong.
- [13] Ohala, J.J. & Riordan, C.J. 1979. Passive vocal tract enlargement during voiced stops. In: J. J. Wolf & D. H. Klatt (eds), *Speech Communication Papers*. New York: Acoust. Soc. of Am., 89 - 92.
- [14] Pompino-Marschall, B. 2004. Von Kempelen's contribution to the theory of acoustic articulation. *Grazer Linguistische Studien* 62, 137-147.
- [15] Reininger, A. 2007. *Wolfgang von Kempelen. Eine Biografie*. Wien: Praesens Verlag.
- [16] Schröder, M. 2009. Expressive speech synthesis: Past, present, and possible futures. In: Tao, Tan (eds), *Affective Information Processing*. London: Springer, 111-126.
- [17] Titze, I. 1994. *Principles of Voice Production*. Englewood Cliffs: Prentice-Hall.
- [18] Trouvain, J. & Brackhane, F. 2010. Zur heutigen Bedeutung der Sprechmaschine von Wolfgang von Kempelen. In: Hoffmann, R. (ed) *Proc. 20. Konferenz Elektronische Sprachsignalverarbeitung Vol. 2 (ESSV)*, Dresden, 97-107.
- [19] von Kempelen, W. 1791. *Wolfgang von Kempelen k. k. wirklichen Hofraths Mechanismus der menschlichen Sprache, nebst der Beschreibung seiner sprechenden Maschine*. Wien: J.V. Degen. Facsimile print of 1970, Stuttgart: Frommann-Holzboog.
- [20] Wokurek, W. & Madsack, A. 2008. Messung subglottaler Resonanzen mit Beschleunigungssensoren. *Fortschritte der Akustik – Proc. DAGA 2008*, Dresden, 125-126.

¹ All citations in this paper were translated by the authors. The original languages of the cited works are German [19], Latin [10] and French [5].